Published in final edited form as:

Learn Individ Differ. 2017 April ; 55: 120–129. doi:10.1016/j.lindif.2017.03.011.

Agreement Among Traditional and RTI-based Definitions of Reading-Related Learning Disability with Preschool Children

Trelani F. Milburn, Ph.D.,

Florida Center for Reading Research, Florida State University

Christopher J. Lonigan, Ph.D.,

Department of Psychology and Florida Center for Reading Research, Florida State University

Darcey M. Allan, Ph.D., and

Department of Psychology, Florida State University (now at Department of Psychology, Ohio University)

Beth M. Phillips, Ph.D.

Department of Educational Psychology and Learning Systems and Florida Center for Reading Research, Florida State University

Abstract

To investigate approaches for identifying young children who may be at risk for later reading-related learning disabilities, this study compared the use of four contemporary methods of indexing learning disability (LD) with older children (i.e., IQ-achievement discrepancy, low achievement, low growth, and dual-discrepancy) to determine risk status with a large sample of 1,011 preschoolers. These children were classified as at risk or not using each method across three early-literacy skills (i.e., language, phonological awareness, print knowledge) and at three levels of severity (i.e., 5th, 10th, 25th percentiles). Chance-corrected affected-status agreement (CCASA) indicated poor agreement among methods with rates of agreement generally decreasing with greater levels of severity for both single- and two-measure classification, and agreement rates were lower for two-measure classification than for single-measure classification. These low rates of agreement between conventional methods of identifying children at risk for LD represent a significant impediment for identification and intervention for young children considered at-risk.

Keywords

learning disability; response to intervention; identification; preschool

Correspondence regarding this article should be directed to Christopher J. Lonigan, Department of Psychology, Florida State University, 1107 W. Call Street, Tallahassee FL 32306-4301 (lonigan@psy.fsu.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Young children are considered at risk for reading-related learning disabilities when, despite high quality instruction, they have difficulty acquiring the foundational competencies involved in reading and writing. For these children, the achievement gap that exists between their early literacy knowledge and skills and those of their typically developing classmates is expected to widen, resulting in increasing difficulty for these children to benefit from classroom instruction. Furthermore, children experiencing academic difficulties are likely to experience concomitant lowered self-esteem, reduced motivation for academic activities and learning, and frustration that has been associated with inattentive and disruptive behavior (Arnold & Doctoroff, 2003). Studies have indicated that this downward spiral may be avoided for children who are identified early and who are provided with intervention at the appropriate level of intensity to meet their learning needs (Torgesen, 2000). However, identification has focused largely on school-aged children despite recent recommendations to avoid a period of "waiting to fail" for early identification of young children who are at risk (Al Otaiba et al., 2014; O'Connor, Bocian, Beebe-Frankenberger, & Linklater, 2010) with some researchers advocating for identification of children prior to kindergarten (Vellutino, Scanlon, Zhang, & Schatschneider, 2008). What is unclear is whether the same methods of indexing learning disabilities with older children can be used to index risk for reading-related LD for preschool-age children.

The reauthorization of the Individuals with Disabilities Education Act (IDEA, 2004) brought about significant changes to regulations for the identification of LD in school-aged children, such as, no longer requiring a severe discrepancy between children's intellectual ability and academic achievement as well as the inclusion of a process that considers children's response to research-based interventions. However, there remain issues that make classifying children as having LD difficult (Fletcher, Lyon, Fuchs, & Barnes, 2007). First, LD is an unobservable construct that has been operationalized by a number of definitional frameworks, and the use of these different methods to index LD may result in inconsistencies in the children identified, depending on how the construct is measured (Compton, Fuchs, Fuchs, & Bryant, 2006; Fuchs & Deshler, 2007; McMaster, Fuchs, Fuchs, & Compton, 2005; Speece, Case, & Molloy, 2003). Second, children's achievement scores on most population-based measures exist on a normal distribution, which requires the use of cut-off scores to identify those children who fall below this score as LD and those above the score as not LD (Francis et al., 2005; Torgesen, 2000; Vaughn, Linan-Thompson, & Hickman, 2003). Given continued efforts to reduce the prevalence of LD coupled with a downward extension to identify risk with children as young as preschool-age, it is necessary to gain a better understanding of the extent to which there is agreement between different methods of indexing LD when used to identify risk with a younger sample.

1.1. Methods of Indexing Learning Disability

One of the common elements across most methods of indexing LD is the concept of unexpected low achievement. In large part, definitional frameworks differ in how the "unexpected" part of this concept is operationalized. The IQ-achievement discrepancy method of indexing LD, historically the most widely recognized and commonly used

method of LD classification, is defined as low achievement in combination with a discrepancy between a child's IQ score and scores on achievement measures. This definitional framework operationalizes the unexpected concept as a deviation from the achievement expected on the basis of IQ (Burt, 1950; Fletcher, Morris, & Lyon, 2003), and children identified with LD using this method are considered to be qualitatively different from other low-achieving students (Rutter & Yule, 1975). However, the validity of the IQ-achievement discrepancy method of identifying LD has been challenged on several fronts. Evidence that a distinct group of underachievers exists is mixed, with most researchers failing to find a bimodal distribution of poor performers (Rodgers, 1983; Share & Silva, 1987) and instead finding a continuous distribution of scores (Francis et al., 2005). Qualitative differences are not found between low-achieving children with and without discrepant IQ scores (Fletcher, Francis, Morris, & Lyon, 2005; Hoskyn & Swanson, 2000; Stuebing et al., 2002). Moreover, children with low achievement in reading, both with and without IQ-achievement discrepancy, have similar cognitive and phonological deficits (e.g., Fletcher et al. 1994; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Hoskyn & Swanson, 2000; Siegel, 1989), and both groups demonstrate similar levels of improvement as a result of intervention (e.g., Vellutino, Scanlon, & Lyon, 2000). A concern with the use of this method to index risk for LD, in addition to questions about validity, is that the requirement of an IQ-achievement discrepancy may arbitrarily exclude a subgroup of low-achieving children who may benefit from intervention (Vaughn & Fuchs, 2003) and for whom a delay in intervention may result in the academic difficulties becoming resistant to remediation (Torgesen et al., 2001).

The Response to Intervention (RTI) approach to indexing LD has garnered support over the past decade as a viable alternative to the traditional IQ-achievement discrepancy model (U.S. Department of Education, 2004). This approach defines LD as failure to make adequate progress, as established by classroom, school, regional, or other expectations of progress, despite comprehensive instruction (Fuchs & Fuchs, 2006). The RTI model involves universal screening and monitoring of children considered at risk for poor progress, and providing increasingly intensive and individualized instruction based on children's responsiveness to instruction. Methods of indexing LD that use this framework and that could be applied to younger children include (a) identifying children who demonstrate low achievement (based on a one-time assessment) after receiving comprehensive instruction (Good, Simmons, & Kame'enui, 2001; Torgesen et al., 2001), (b) identifying children who demonstrate insufficient or low growth in the index skill (Fuchs, Fuchs, & Compton, 2004; Vellutino et al., 1996), and (c) identifying children who demonstrate both low achievement and low growth, referred to as dual discrepancy (Fuchs & Fuchs, 1998; Speece & Case, 2001). These methods are currently used to identify children considered at risk based on RTI universal screening results and monitoring results, but little is known about the agreement between the methods to identify children who are at risk.

1.2. Agreement Among Methods of Indexing LD with Older School-aged Children

Although there has been much discussion in the literature regarding the comparative advantages and disadvantages of specific methods of indexing LD (Fletcher et al., 2007; Fuchs & Deshler, 2007; Fuchs & Fuchs, 2006; O'Connor & Klingner, 2010; Tran, Sanchez,

Arellano, & Swanson, 2011), there has been relatively less discussion regarding the extent to which these different methods result in overlap in the individuals identified. In studies of college-age students, researchers have reported that although many of the different LD classification approaches resulted in similar numbers of students being identified as having LD, the agreement between methods to classify LD was generally modest (Brackett & McPherson, 1996; Proctor & Prevatt, 2003).

Recent research examining agreement among methods of indexing risk for LD with young school-age children has also resulted in modest overlap in the children identified. Barth et al. (2008) examined agreement among three methods of indexing risk for LD within the RTI framework (i.e., low achievement, low growth, dual-discrepancy) to classify first-grade children as adequate or inadequate readers using three cut-off points (i.e., 0.5, 1.0, and 1.5 standard deviations below the mean) and different types of measures (i.e., norm-referenced tests, curriculum-based tests). The authors reported moderate agreement of the children classified using each method across cut-off points. Brown-Waesche, Schatschneider, Maner, Ahmed, and Wagner (2011) examined agreement among traditional IQ-achievement discrepancy and RTI-based methods of indexing LD with a large sample of children in grades one to three. Using six cut-off points (i.e., 3rd, 5th, 10th, 15th, 20th, 25th percentiles) for the Word-Reading Fluency and Non-Word Fluency subtests from the Dynamic Indicators of Basic Early Literacy Skills (Good, Kaminski, Smith, Laimon, & Dill, 2001), they reported generally poor agreement between methods across the range of cut-off points. Thus, in both adult and school-age samples, the individuals identified as LD using one classification method are often not identified as LD using a different classification method.

1.3. Indexing Risk Status During Preschool

Understanding academic-related deficits that emerge in the early years is critical given the strong evidence of the stability of these skills across early- and middle-childhood (Duncan et al., 2007; Hogan, Catts, & Little, 2005; Lonigan, Schatschneider, & Westberg, 2008; Muter, Hulme, Snowling, & Stevenson, 2004). For example, research has shown that early letter knowledge and phonological awareness are strongly linked to later decoding skills (Lonigan et al., 2008) and early oral language is predictive of later reading comprehension skills (Butler, Marsh, Sheppard, & Sheppard, 1985; Storch & Whitehurst, 2002). However, it is necessary to determine whether the children identified as at-risk and not-at-risk using these emergent literacy skills are the same or different depending on the definitional framework used for classification. High levels of disagreement of the children identified as at risk for LD using different methods would indicate that different children would be eligible for services and special accommodations contingent on which method was used to determine risk status. To avoid over-identification or false positives of children who are at risk, an outcome that would unduly stress limited resources to provide monitoring or intervention for children who really need it (Fletcher et al., 2002; Jenkins & O'Connor, 2002), the use of multiple measures to assess the same early literacy skill has been recommended (Compton et al., 2010; Fletcher, Francis et al., 2005; Francis et al., 2005). Additionally, evaluating how different factors (e.g., deficit severity, deficit pervasiveness) influence agreement levels may provide insight to the benefits and drawbacks of using a particular method of indexing risk (Brown-Waesche et al., 2011).

1.4. Current Study

The goals of this study were to identify agreement among four methods used to index LD with older children (i.e., IQ-achievement discrepancy, low achievement, low growth, dual-discrepancy) to index risk for LD with a large sample of preschool-age children. The first question of this study concerned the level of agreement between paired comparisons of methods to index risk for LD using single measures of early literacy skills (i.e., language, print knowledge, phonological awareness) across three levels of severity (i.e., 25th, 10th, 5th percentiles). Based on the findings of Brown-Waesche et al. (2011), it was expected that agreement between methods of classification for risk would be weak-to-moderate across the early literacy skills and that rates of agreement would improve as the levels of severity required for this classification change from the most to the least stringent criteria. The second question of this study was whether the use of multiple measures to determine classification of risk for LD improved agreement over the use of single measures. Based on recommendations by Compton et al. (2010) and Fletcher, Francis et al. (2005), it was expected that agreement among different methods of indexing risk for LD would be stronger when LD classification required pervasive deficits (i.e., deficits on two measures of a construct) than it would be when LD classification only required deficits on a single measure. The primary metric used was the Chance-Corrected Affected Status Agreement (CCASA; Brown-Waesche et al., 2011). This statistic was preferred over other methods of quantifying agreement levels (e.g., Kappa) because the questions of the current study focused only on agreement levels for children identified as at risk for LD, and CCASA does not include children not identified as at risk for LD in the calculation.

2. Method

2.1. Participants

Participants in this study included 1,011 preschool children who were part of a longitudinal study designed to examine development of children at risk of poor academic achievement. The children were recruited from 43 different preschools that were either part of the local school district's Title I preschool program ($n = 21$) or private fee-for-service preschools serving the same community ($n = 22$). Children were recruited in three cohorts across three successive school years. There were 30 preschools in Year 1 (i.e., 14 Title I, 16 private), 36 preschools in Year 2 (i.e., 21 Title I, 15 private), and 31 preschools in Year 3 (i.e., 18 Title I, 13 private). The children ranged in age from 53 to 70 months ($M = 61.48$, $SD = 3.64$). Children in the sample were predominantly white or black/African American (i.e., 47.6% white, 42.0% black/African American, 3.0% Hispanic/Latino, 2.3% Asian, 0.3% American Indian, 3.5% Mixed Race, 1.3% unknown), and 45% were female.

The larger project for this study focused on children with higher than average risk for later educational difficulties. Because of admissions requirements, all children from the Title I preschools whose parents returned consents were eligible for inclusion in the study. In contrast, the private preschools had an open enrollment policy, and comparison of children across type of center in Year 1 indicated that children from the private centers had substantially higher scores at preschool entry than children from the Title I preschools. Therefore, in Years 2 and 3, children from the private preschool were screened using the 25-

item Revised Get Ready to Read! screening tool (R-GRTR; Lonigan & Wilson, 2008) and were excluded if their scores were 21 or higher. Across both types of centers, children were excluded if there was evidence of severe developmental, neurological, or sensory impairment, as reported by the educator or based on observation of the child by the examiners.

2.2. Measures

Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP; Lonigan, Wagner, Torgesen, & Rashotte, 2002): Four subtests of the Pre-CTOPPP, which was the development version of the Test of Preschool Early Literacy (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007), were administered, including Definitional Vocabulary, Blending, Elision, and Print Knowledge. The Pre-CTOPPP Definitional Vocabulary subtest elicits two responses for each item. For the first part of each item, the child is asked to label a single image or group of images and for the second part of each item, the child is asked to respond to a follow-up question regarding function or relevant context for the item. The Blending subtest requires children to blend components of words spoken by the examiner to form a new word (e.g., blending “star” and “fish” to form “starfish,” blending /f/ and “ox” to form “fox”). The Elision subtest requires children to remove a part of a word spoken by the examiner to form a new word (e.g., removing “snow” from “snowshoe” to create “shoe,” removing /d/ from “raid” to create “ray”). All answers on the Blending and Elision subtests are real words, and items spanned the range of linguistic complexity (e.g., compound words to phonemes). Each subtest includes nine multiple-choice items, which required children to point to one picture out of four that represented the correct response, and free-response items, which required children to verbally produce the correct response. The Print Knowledge subtest requires children to respond to questions related to print concepts (e.g., “Which one is a letter?”, “Which one can you read?”), letter-name recognition, letter-sound recognition, and letter-name and letter-sound production. All Pre-CTOPPP subtests include at least one practice item and all items were administered to all children without ceiling criteria. These four subtests have good reliabilities for 3- to 5-year-olds (i.e., α s = .86 to .96) and substantial evidence of validity (e.g., concurrent validity coefficients of .59 to .77 with other measures of similar constructs).

Clinical Evaluation of Language Fundamentals-Preschool (CELF-P; Wiig, Secord, & Semel, 1992): The Expressive Language composite score of the CELF-P comprises subscale scores from three subtests including the Recalling Sentences in Context subtest, which examines the ability to recall and repeat sentences, the Formulating Labels subtest, which examines the ability to name pictures that represent nouns and verbs, and the Word Structure subtest, which examines knowledge of morphological rules and forms. According to the test manual, the internal consistency of the Expressive Language composite ranges from .90 to .95 for this age group, and the composite score has good test-retest reliability ($r = .92$).

Test of Early Reading Ability-3 (TERA-3; Reid, Hresko, & Hammill, 2001): The Alphabet subtest of the TERA-3 was used as a measure of print knowledge. This subtest measures children’s knowledge of the alphabet and sound-letter correspondence. According

to the test manual, the internal consistency of the TERA-3 Alphabet subscale ranges from .92 to .94 for this age group, and test-retest reliability with preschoolers is good ($r = .94$).

Cognitive abilities measures: The Block Design and Matrix Reasoning subtests of the Wechsler Preschool and Primary Scales of Intelligence, 3rd Edition (WPPSI; Weschler, 2002) were administered to children to obtain a measure of general cognitive abilities. On the Block Design subtest, children are required to replicate increasingly complex geometric patterns from either a picture or a model, with a limited amount of time for each pattern to be replicated. On the Matrix Reasoning subtest, children are required to identify the one stimulus, out of four, that completes the two-by-two matrix of stimuli. Both subtests have good reliability, and they have significant correlations with Full Scale IQ scores derived from the WPPSI-III.

2.3. Procedure

The protocol for this study was approved following full committee review by the (name redacted for review) Institutional Review Board. Prior to data collection, informed consent/permission was obtained from children's parents or guardians. Children were administered the battery of assessments by trained research assistants in quiet areas of the children's preschools. All research assistants had received training in the administration of these measures and had demonstrated mastery of the testing protocols in one-on-one sessions with the project's assessment coordinator. All assessments used in this study were administered to the children at three time points during the preschool year (i.e., fall, winter, spring). At each time point, the children provided assent to participate in the study. The assessments were completed over three to four 20- to 30-minute sessions within a two-week period. During testing, children were given breaks if requested or if the examiner noticed fatigue or distraction. Order of test administration varied across children. Upon completion of testing, the children were given a choice of stickers, a pencil, or a book. All measures were scored independently by two research assistants to ensure accurate scoring procedures, and disagreements were resolved by discussion and consultation with the assessment coordinator.

2.3.1. Method of Indexing Risk Status—Six measures were used to represent the three distinct early literacy skill domains, including language (i.e., Pre-CTOPPP Definitional Vocabulary, CELF Expressive Language), print knowledge (i.e., Pre-CTOPPP Print Knowledge, TERA Alphabet), and phonological awareness (i.e., Pre-CTOPPP Elision, Pre-CTOPPP Blending). Given concerns regarding what percentile cut-point is appropriate for determining the presence of LD (L. S. Fuchs and Fuchs, 1998; Speece & Case, 2001), each method of indexing risk for LD was applied at three levels of severity (i.e., 25th, 10th, & 5th percentiles). Standard scores on three of the measures (i.e., Pre-CTOPPP Elision, Pre-CTOPPP Blending, TERA Alphabet) were scaled scores, which were converted to standard score equivalents (i.e., $M = 100$; $SD = 15$) to make scores comparable across measures. Children were classified as at risk for LD or not at risk for LD based on the four methods:

Unexpected low achievement (ULA) classification: Children were classified as at risk for LD using ULA if they (a) demonstrated a one standard deviation (i.e., 15 point) discrepancy

between their average scores on IQ measures and scores on an end-of-year emergent literacy skill measure, and (b) scored below the specified percentile cut-point on the same achievement measure. To create a proxy measure of IQ, scores on the Block Design and the Matrix Reasoning subtests from the WPPSI were averaged and converted to a standard score equivalent to make it comparable to the standard scores on the achievement measures. Children were excluded from ULA analysis for a particular measure if they did not have scores on that measure at the end of the preschool year or did not complete the WPPSI subtests. Across measures, data were available for between 1,001 and 1,006 students.

Low achievement (LA) classification: Children were classified as at risk for LD based on LA if their scores at the end of the year were below the specified percentile cut-point (i.e., 5th, 10th, 25th percentiles). Children were excluded from analyses for a measure if they did not have data for that measure at the end of preschool. Across measures, data were available for 1,001 to 1,006 students.

Low growth (LG) classification: Children were classified as at risk for LD based on their growth in skills over the preschool year (i.e., fall, winter, spring). Latent-growth-curve models were constructed using Mplus version 5.2 (Muthen & Muthen, 1998) to estimate slopes of growth separately for each measure. Slope estimates were converted to *z*-scores and children were classified as at risk for LD if their *z*-score fell below the *z*-score corresponding to the specified percentile rank. Students were excluded from the analysis for a particular measure if they were missing scores for more than one of the three time-points. Across measures, growth data were available for between 998 and 1,004 students.

Dual discrepancy (DD) classification: Children were classified as at risk for LD based on DD if they were classified as both LG and LA at the specified percentile cut-point. Across measures, data were available for between 998 and 1,004 students.

2.3.2 Agreement—Once classifications for risk status were made separately for each measure, agreement among all possible pairs of alternative methods was compared using the proportion of children identified by each method and the CCASA statistic. In addition to this classification, a pervasive risk status for LD was examined for each of the four methods of indexing risk for LD. Pervasive risk for LD was defined as meeting the classification criteria on both measures of an early literacy skill domain (i.e., language, print knowledge, phonological awareness). As with the single-measure classification, agreements among all possible pairs of alternative pervasive methods were compared using the proportion of children identified by each method and the CCASA statistic.

This CCASA statistic was chosen over other matrices of agreement (e.g., Kappa) because it measures agreement between methods of classifying individuals who are affected by a target condition (e.g., classification for risk of LD) without inflating agreement by including individuals who are not affected by the condition of interest. Affected Status Agreement (ASA) is calculated using two-by-two crosstabs wherein the number of children identified as affected by both classifications being compared (e.g., ULA and LG) is divided by the total number of children identified as affected by both or either classification. To correct for the amount of agreement expected due to chance, a CCASA is calculated. Initially, the amount

of agreement expected due to chance is calculated for each cell by multiplying the respective base rates of the proportion of children identified for that cell. Then the amount of agreement expected due to chance when the two methods agree on a classification of risk is divided by the sum of the amount of agreement for either method that determined classification of risk. Next, the amount of the affected status agreement due to chance is subtracted from the ASA divided by 1 minus the affected status agreement due to chance to derive the CCASA for that comparison.

3. Results

Descriptive statistics of the children's ability on the six measures of early literacy skills and the two measures of non-verbal IQ are shown in Table 1. Overall, the scores of this sample of children were generally below average levels. The percentages of the sample that met the criteria to be considered at risk for LD by each of the four methods are shown in Table 2 for the six single measures of early literacy at the three levels of severity. All four methods classified children as at risk for LD at all three levels of severity for single measures, with the exception of Pre-CTOPPP Definitional Vocabulary, which did not identify any children at the 5th percentile using the DD method. The percentage of children who met the ULA classification was small relative to the percentages of children who were classified as at risk for LD based on the three RTI methods across early literacy skills at the three levels of severity. The LA method had the highest percentages of children classified as at risk for LD.

3.1. Agreement for Risk Using Single Measures

To address the first question of this study, the CCASA statistics for paired comparisons of the four methods of indexing risk for LD for each of the six single measures of early literacy at each level of severity are shown in Table 3. Most of the agreement values were greater than chance but some were below chance levels (i.e., See bolded confidence levels for paired comparisons that were not above chance). Overall, the ULA method had the lowest levels of agreement with all three RTI methods of indexing risk and CCASA rates of agreement between the RTI methods revealed higher agreement across the three levels of severity compared to those with ULA method. The strongest agreements were between the pairings of DD and LG and also DD and LA methods, likely due to the overlapping criteria of these methods (i.e., DD included both LG and LA).

Across the six single measures of early literacy skills, agreement rates varied for all methods of indexing risk for LD and at all levels of severity. Notably, the Pre-CTOPPP Definitional Vocabulary subtest had the lowest agreement across LD methods of classifying risk, with the exception of the ULA versus LA methods comparison. In fact, the highest rate of agreement for the Pre-CTOPPP Definitional Vocabulary subtest was achieved when comparing ULA and LA at the 5th percentile. Comparison of CCASA among the methods from the RTI approach indicated that, in general, Pre-CTOPPP Blending demonstrated the highest agreement among the six early literacy measures.

Rates of agreement between different methods of indexing risk for LD varied across the three levels of severity (i.e., cutoffs), but, generally, these rates decreased with increasing levels of severity. However, agreement rates between the ULA and LA methods followed a

different pattern whereby rates of agreement increased with increasing levels of severity. There was no consistent pattern of change at different levels of severity for agreement between ULA and LG, ULA and DD, and the LG and DD methods.

3.2. Agreement for Risk Using Multiple Measures

To address the second question of the study, the percentages of children who met criterion when both measures within a skill domain were required for classification of risk across each of the four methods of indexing risk are shown in Table 4. Each of the four methods identified children at risk for LD at all levels of severity, with the exception of the language construct for which no children were classified at the 5th percentile for either LG or DD. Overall, fewer children were classified as at risk for LD across the four methods at all levels of severity when the method to determine risk relied on two measures rather than on just one measure. Percentages of children classified as at risk for LD were generally lower for the ULA method than the RTI methods across the three levels of severity and again, LA identified the highest percentages of children across the three early literacy constructs. Overall, larger percentages of children were classified as at risk for LD with pervasive deficits in the print knowledge domain at the 25th percentile across the four methods but not at the 10th and 5th percentiles.

The CCASA statistics for the four methods of classifying risk at each level of severity for pervasive risk status for LD are shown in Table 5. Overall, rates of agreement were lower for classification of risk for LD based on pervasive deficits compared to the use of single measures. Agreement rates between the ULA method and the RTI methods showed a great deal of variability across the three early literacy constructs with the lowest agreement between ULA and LG. As with the single-measure use, agreement was highest for comparisons of RTI methods with overlapping definitional requirements (i.e., DD with LA or LG). For specific constructs, the CCASA statistics were higher for the language construct between the ULA and the LA methods compared to phonological awareness and print knowledge.

No general patterns emerged in terms of changes in the magnitude of CCASA across level of severity for pervasive risk status, with the exception of the agreement between the ULA and the LA method, which increased with greater severity. CCASA statistics were higher for the language construct between the ULA and the LA methods compared to phonological awareness and print. There was weak agreement between the ULA method of identifying risk for LD and both the LG and the DD methods at all three levels of severity for the language construct, and CCASAs were low for all comparisons of the LA, LG, and DD methods relative to each other for all three early literacy constructs when pervasive deficits were considered.

4. Discussion

The purpose of this study was to examine agreement between paired comparisons of four methods of indexing risk for learning difficulties or LD with children during the preschool year based on methods of indexing LD with older school-aged children. Three core early literacy domains were examined at three levels of severity using both a single-measure

approach and a two-measure approach for classifying children as at risk for LD. Across methods of identifying risk status, measures, and levels of severity, substantial numbers of children in this sample were classified as at risk for LD; however, rates of agreement across methods of classifying risk for LD were weak-to-moderate.

4.1. Identification of Children At Risk for Reading-Related Learning Disabilities

In this study, all four methods of classifying risk for reading-related LD identified preschool children at each of the three levels of severity and across most of the single measures of emergent literacy. Prior studies have examined unexpected low achievement with kindergarten children based on screening or monitoring for responsiveness (e.g., Catts, Nielson, Bridges, Liu, & Bontempo, 2015; O'Connor, Bocian, Sanchez, & Beach, 2012; Vellutino et al., 2008) but only a few studies have extended this identification to preschool children (Lonigan & Phillips, 2016). The current study is unique in that the utility of contemporary definitional frameworks for identifying preschool children at risk for reading-related LD and the levels of agreement among these methods was examined.

The use of the IQ-Achievement discrepancy or ULA method identified the fewest children of all of the LD methods at all levels of severity and across the six measures of early literacy. Notably, the percentage of children identified as at risk for LD using this method at the 25th percentile was comparable to national estimates of 6% (U.S. Department of Education, 2014); however, given the at-risk status of this sample of children, many of whom were from economically disadvantaged families, the percentage of children identified using the ULA method was conservative. That is, a rate higher than the national average would be expected for this sample of children given recommendations for the importance of early identification and intervention (Fletcher, Coulter, Reschly, & Vaughn, 2004). In contrast to the ULA method, the LA method consistently identified the highest percentage of children as at risk for LD across measures and levels of severity. As the most inclusive method, LA identified 100% of the children with low achievement with an IQ discrepancy (ULA) as well as additional children who scored below the cutoff points in achievement but also scored low in cognitive ability. Given results of previous studies, which have demonstrated that both groups benefit when provided with adequate instruction (Foorman, Francis, & Fletcher, 1995; Vellutino, et al., 2000), an inclusive method may be more desirable than a conservative one to avoid false negatives, or children not identified as at-risk when they actually are at-risk at this early point in literacy instruction for children who are at risk (Torgesen, 2002).

4.2. Agreement Among Methods of Determining Risk Status

Rates of agreement for determining risk for reading-related LD with preschool children varied widely among the methods for classification in this sample. However, our rates of agreement for the 162 comparisons conducted in this study (i.e., 108 for single measures and 54 for pervasive deficits) using CCASA indicated that 21% of these comparisons (i.e., 34/162) resulted in agreement above .40, whereas Barth et al. (2008) reported only 15% of their kappas were above .40, and a substantial proportion of this agreement using kappas included individuals who were not poor readers or at-risk for a reading-related LD. Brown-Waesche et al. (2011) reported that 59% or 164/276 comparisons resulted in agreement

higher than .40 using CCASA, possibly reflecting greater stability between LD identification methods for older children (i.e., first to third grade). On the whole, these rates of agreement were low, indicating that a large number of different children were identified by the different methods of classification..

4.2.1 Single-measure classification—There was a wide range of agreement between the different methods of determining risk for reading-related LD using single measures of early literacy. Agreement levels were generally lowest between the ULA and the RTI methods of determining risk for LD indicating that these approaches to classification of risk for LD were not identifying the same children. Comparatively, rates of agreement between RTI approaches were higher. The highest rates of agreement were between the DD and LA methods and the DD and LG methods; however, these methods have overlapping definitional requirements for classification of risk that, in part, accounts for higher levels of agreement. One distinct comparison within the RTI approaches is the LA and LG methods of indexing risk that, apart from Pre-CTOPPP Definitional Vocabulary, had comparable rates of agreement across measures and levels of severity to the other RTI comparisons. This finding indicates that, rather than simply casting a wide net to identify the largest number of children who may be at-risk, as might be the case with LA given high rates of classification of risk for LD, growth may be a more consistent means of identifying preschool children's risk status based on these higher agreement rates. However, examination of the data indicated that some of the children identified as at risk for LD using the LG method had high scores at Time 1 and Time 3 resulting in false positives (i.e., incorrectly identified as at-risk) due to low growth over time when the low growth was, in part, a function of having higher initial scores and, thus, less room to grow in that skill area. Alternatively, the DD method requires both low growth and low achievement and may serve to avoid incorrectly identifying children at risk for LD because children whose low growth is a function of the fact that they started with high scores are not identified as at risk for LD.

There continue to be questions about the validity of growth over time compared to a static measure for defining risk for LD. For instance, Schatschneider, Wagner, and Crawford (2008) conducted a large-scale longitudinal study of the relative contributions of growth in oral reading fluency as measured at four time points during first grade and a static score at the end of first grade for predicting concurrent and end of second grade reading achievement. The results indicated that the growth in oral reading fluency did not explain unique variance in second grade reading achievement beyond that of a single measure of oral reading fluency at the end of first grade. Brown-Waesche et al. (2011) also reported that growth was a less reliable measure compared to a static measure of low achievement. Other studies, however, have indicated that low achievement lacked the sensitivity necessary to correctly identify those children who had reading-related LD and proposed that dual-discrepancy is a more reliable method for classification of LD than static achievement or 'growth rate alone' approaches (e.g., McMaster et al., 2005; Speece & Case, 2001; Speece et al., 2003). These studies, together with the findings of the current study, demonstrate that higher agreement in classification of LD across definitions include growth. More work is needed to determine the validity of different definitional frameworks to classify preschool children as at risk for LD using measures of early literacy.

4.2.2. Two-measure classification—The use of the two-measure approach (i.e., pervasive) to identify children as at risk for reading-related LD was expected to corroborate a finding of deficit within and between methods; however, there was poorer agreement when two measures were used than when single measures were used. Although more information is expected to result in greater reliability of the LD classification (Fletcher, Denton, & Francis, 2005), in the current study, there was not greater agreement across methods for classification of risk for LD using two measures. There were considerable differences in rates of agreement for some single measures within a construct (e.g., Pre-CTOPPP Definitional Vocabulary and CELF Expressive Language) indicating that the two measures were not measuring the same underlying construct. Selection of measures for identification of risk status for LD relies on the use of tests that measure the same construct.

Similar to the use of single measures, the ULA method generally identified fewer children in a two-measure approach and the LA method identified the most children. In terms of individual emergent literacy constructs, generally, the most children were identified across levels of severity for the print knowledge construct, although the numbers were low at the highest levels of severity. The agreement rates between the traditional IQ-achievement discrepancy (i.e., ULA) and the RTI methods were generally poor across the three emergent literacy constructs. The largest levels of agreement were for comparisons of RTI methods with overlapping definitional requirements, with the strongest agreements between DD and LG. No general patterns emerged in terms of changes in the magnitude of agreement across levels of severity with the exception of the agreement between ULA and LA that increased with greater severity. Again, at increasing levels of severity, children who demonstrated low achievement as measured by deficits on two measures were more likely to also demonstrate a discrepancy between scores on achievement and cognitive measure.

4.3. Level of Agreement Across Severity

To examine differences in levels of severity of deficit across methods of classification of risk for LD, the current study compared three cutoff points. First, as might be expected, fewer children were identified as LD at increasing levels of severity. Second, the rates of agreement between different LD methods generally decreased with increasing level of severity. That is, the rate of agreement typically followed a pattern of lower agreement from the 25th to the 10th to the 5th percentiles. This finding is consistent with Brown-Waesche et al. (2011) who reported a similar pattern of reduced agreement with greater severity of deficit. This pattern can be explained by the use of the CCASA statistic that includes only children identified as at risk by one or both methods being compared. Furthermore, when the criteria for classification is less stringent (e.g., 25th percentile), larger proportions of children's scores are distributed farther from the cutoff increasing the likelihood that there would be agreement on, for example, not achieving either the level of performance or growth. For children whose scores are below the more stringent level (i.e., 5th percentile), greater proportions of these children are likely to have true scores just below the cutoff, resulting in these children changing classification of at risk and not at risk with repeated measurement and across measures due to measurement error (Francis et al., 2005). The one exception to this pattern of decreasing agreement with increasing severity of deficit was found for agreement between the ULA and LA methods. Agreement rates between these two

methods increased with increased severity, a finding most likely attributable to an overlapping definitional requirement of low achievement scores for both methods.

4.4. Level of Agreement Across Measures

Examination of rates of agreement across the six measures of early literacy revealed several reliable differences in agreement. First, the CELF-P Expressive Language scale generally had the highest levels of agreement of the six measures across all levels of severity for both the IQ-achievement discrepancy comparisons and comparison of methods within the RTI approach (i.e., across all comparisons and levels of severity ranging from .15 to .89). Because scores for this measure are based on a composite of subtests, the higher rates of agreement may be attributable to the higher reliability obtained by averaging multiple related measures. Across all comparisons, Pre-CTOPPP Blending generally had the second highest rates of agreement and this measure maintained the highest rates of agreement across all levels of severity in methods that included growth. Although letter naming has been used in studies as the screening measure to identify kindergarten children at risk for LD (Torgeson et al., 1999; Vellutino et al., 2008) and has been found to be a significant predictor in young children's reading outcomes (Catts, Fey, Zhang, & Tomblin, 2001; Catts et al., 2015; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004), none of the single measures of print knowledge yielded consistently high levels of agreement relative to the other measures.

4.5. Implications

Children in this study were attending preschools where they received varying degrees of exposure to high-quality instruction that targeted language, phonological awareness, and print knowledge, and some children were exposed to additional instruction--similar to Tier II instruction in an RTI model. Consequently, the fact that sizable numbers of children were classified as at risk across methods of indexing risk indicates that identification at preschool-age is feasible and even necessary to avoid a "Wait-to-Fail" approach that might delay intervention for children who are at risk for reading-related LD (Al Otaiba et al., 2014; O'Connor et al., 2010). Given that this study identified comparable levels of agreement between methods of classifying risk with preschool-age children and the results reported in studies with older children (e.g., Barth et al., 2008), the weak agreement found herein was not related to the young age of the children but to inherent problems in the definitional frameworks used to classify risk status.

In applied settings, available finances and resources play a part in the selection of methods for indexing children's risk status and the level of severity of the criteria used, in part because of the likelihood of different definitional frameworks to identify larger or smaller numbers of children. However, the prospect of over-identifying children (i.e., false positives) needs to be weighed against the risk of large numbers of false negatives and the challenge of providing remediation for those children at a later point, given that they may become more resistant to intervention (Torgeson, 2000). In the current study, the highest levels of agreement were between RTI methods; however, due to overlapping requirements for classification of risk between DD and its constituent parts (i.e., LA and LG), it may be advisable to use a DD model until such time as further research can establish the validity of

the use of one or both of these methods with this young population. Additionally, despite variability in agreement across single measures of early literacy, there was no consistent pattern favoring one measure over another in terms of rates of agreement, and, unexpectedly, levels of agreement between methods to determine risk for LD were lower when using the two-measure classification than when using the single-measure classification. These results should not be taken as promoting the use of single measures over multiple measures; rather, they highlight the importance of selection of tests that measure the same underlying construct. Finally, the generally low agreement among methods of classification highlights the need for future research to evaluate the pragmatic utility of the different approaches to achieve specified goals of early identification.

4.6 Limitations

Despite the strengths of this study, including a relatively large sample, the use of multiple measures to index constructs, and the comparison of multiple methods of identifying risk for LD, there are a number of limitations worth noting. First, the children who participated in this study were considered at high risk for poor academic achievement due to economic disadvantage, and the rates of identification and agreement between methods for classification of risk of LD may be unique to this population. Second, the children's responsiveness to instruction was based on instruction in their preschool program over the year and not necessarily more intensive intervention. Third, growth was measured at three time points and not more frequently as might be the case in progress monitoring. Finally, this study examined agreement between methods for classification of LD with preschool children at an early point in their literacy instruction. Studies are needed to investigate the longitudinal stability of the use of these methods for classification of LD in preschool as children progress through conventional literacy instruction.

4.7. Conclusion

LD is an unobservable construct that has been operationalized by a number of definitional frameworks to determine children's risk status for LD through screening or monitoring responsiveness to instruction or intervention. The findings of this study identify inherent problems with the current utility of different methods to index risk, cutoff points at different levels of severity, and the use of single versus multiple measures to classify young children as at risk for LD. Ultimately, different children are identified as at risk across different methods of indexing risk status representing a significant impediment for achieving the goals of early identification. Moreover, this disparity in identification of the correct children extends to our ability to confirm the effectiveness of early intervention and for whom it is effective given the very real possibility that, with the wide range of methods and criteria used, samples include high numbers of false positives and worse, do not include high numbers of children that truly are at-risk.

Acknowledgments

This research and report was supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (HD052120 & HD060292) and the Institute of Educational Science (R305F100027 & R305B090021). The views expressed herein are those of the authors and have not been reviewed or approved by the granting agency.

References

- Al Otaiba S, Connor CM, Folsom JS, Wanzek J, Greulich L, Schatschneider C, Wagner RK. To wait in Tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children*. 2014; 81:11–27. [PubMed: 25530622]
- Arnold DH, Doctoroff GL. The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*. 2003; 54:517–545.
- Barth AE, Stuebing KK, Anthony JL, Denton CA, Mathes PG, Fletcher JM, Francis DJ. Agreement among response to intervention criteria for identifying responder status. *Learning and Individual Differences*. 2008; 18:296–307. [PubMed: 19081758]
- Brackett, J., McPherson, A. Learning disabilities diagnosis in postsecondary students: A comparison of discrepancy-based diagnostic models. In: Gregg, N.Hoy, C., Gay, AF., editors. *Adults with learning disabilities: Theoretical and practical perspectives*. New York, NY: Guilford Press; 1996. p. 68-84.
- Brown-Waesche JS, Schatschneider C, Maner J, Ahmed Y, Wagner R. Examining agreement and longitudinal stability among traditional and RTI-based definitions of reading disability using the affected-status agreement statistic. *Journal of Learning Disabilities*. 2011; 44:296–307. [PubMed: 21252372]
- Burt, C. *The backward child*. 3. London, U.K: University of London Press; 1950.
- Butler SR, Marsh HW, Sheppard MJ, Sheppard JL. Seven-year longitudinal study of the early prediction of reading achievement. *Journal of Educational Psychology*. 1985; 77:349–361.
- Catts HW, Fey ME, Zhang X, Tomblin JB. Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*. 2001; 32:38–50.
- Catts HW, Nielsen DC, Bridges MS, Liu YS, Bontempo DE. Early identification of reading disabilities within an RTI framework. *Journal of Learning Disabilities*. 2015; 48:281–297. [PubMed: 23945079]
- Compton DL, Fuchs D, Fuchs LS, Bryant JD. Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*. 2006; 98:394–409.
- Compton DL, Fuchs D, Fuchs LS, Bouton B, Gilbert JK, Barquero LA, ... Crouch RC. Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*. 2010; 102:327–341. [PubMed: 20689725]
- Duncan GJ, Dowsett CJ, Claessens A, Magnuson K, Huston AC, Klebanov P, ... Japel C. School readiness and later achievement. *Developmental Psychology*. 2007; 43:1428–1446. [PubMed: 18020822]
- Fletcher JM, Coulter WA, Reschly DJ, Vaughn S. Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia*. 2004; 54:304–331. [PubMed: 15741940]
- Fletcher JM, Denton C, Francis DJ. Validity of alternative approaches for the identification of learning disabilities: Operationalizing unexpected underachievement. *Journal of Learning Disabilities*. 2005; 38:545–552. [PubMed: 16392697]
- Fletcher JM, Francis DJ, Morris RD, Lyon GR. Evidence-based assessment of learning disabilities in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:506–522. [PubMed: 16083392]
- Fletcher, JM., Lyon, GR., Fuchs, LS., Barnes, MA. *Learning disabilities: From identification to intervention*. New York, NY: Guilford Press; 2007.
- Fletcher, JM., Morris, RD., Lyon, GR. Classification and definition of learning disabilities: An integrative perspective. In: Swanson, HL.Harris, KR., Graham, S., editors. *Handbook of learning disabilities*. New York, NY: Guilford Press; 2003. p. 30-56.
- Fletcher JM, Shaywitz SE, Shankweiler D, Katz L, Liberman IY, Steubing KK, ... Shaywitz BA. Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology*. 1994; 86:6–23.

- Fletcher, JM., Lyon, GR., Barnes, M., Stuebing, KK., Francis, DJ., Olson, RK., ... Shaywitz, BA. Classification of learning disabilities: An evidence-based evaluation. In: Bradley, R.Danielson, L., Hallahan, DP., editors. Identification of learning disabilities: Research to practice. Mahwah, NJ: Erlbaum; 2002. p. 185-250.
- Foorman, BR., Francis, DJ., Fletcher, JM. Growth of phonological processing skill in beginning reading: The lag versus deficit model revisited. Paper presented at the Society for Research on Child Development; Indianapolis, IN. 1995.
- Francis DJ, Fletcher JM, Stuebing KK, Lyon GR, Shaywitz BA, Shaywitz SE. Psychometric approaches to the identification of LD IQ and achievement scores are not sufficient. *Journal of Learning Disabilities*. 2005; 38:98–108. [PubMed: 15813593]
- Francis DJ, Shaywitz SE, Stuebing KK, Shaywitz BA, Fletcher JM. Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*. 1996; 88:3–17.
- Fuchs D, Deshler DD. What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice*. 2007; 22:129–136.
- Fuchs D, Fuchs LS. Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*. 2006; 41:93–99.
- Fuchs LS, Fuchs D. Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice*. 1998; 13:204–219.
- Fuchs D, Fuchs LS, Compton DL. Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly*. 2004; 27:216–227.
- Good, RH., Kaminski, RA., Smith, S., Laimon, D., Dill, S. Dynamic Indicators of Basic Early Literacy Skills-Fifth Edition. Eugene: University of Oregon; 2001.
- Good RH, Simmons DC, Kame'enui EJ. The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*. 2001; 5:257–288.
- Hogan TP, Catts HW, Little TD. The Relationship Between Phonological Awareness and Reading Implications for the Assessment of Phonological Awareness. *Language, Speech, and Hearing Services in Schools*. 2005; 36:285–293.
- Hoskyn M, Swanson HL. Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *School Psychology Review*. 2000; 29:102–119.
- Jenkins, JR., O'Connor, RE. Early identification and intervention for young children with reading/learning disabilities. In: Bradley, R.Danielson, L., Hallahan, DP., editors. Identification of learning disabilities: Research to practice. Mahwah, NJ: Erlbaum; 2002. p. 99-149.
- Lonigan CL, Phillips BM. Response to instruction in preschool: Results of two randomized studies with children at significant risk of reading difficulties. *Journal of Educational Psychology*. 2016; 108:114–129. [PubMed: 26869730]
- Lonigan, CJ., Schatschneider, C., Westberg, L. Developing early literacy: Report of the national early literacy panel. Washington, DC: National Institute for Literacy; 2008. Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling; p. 55-106.
- Lonigan, C., Wagner, R., Torgeson, J., Rashotte, C. Preschool comprehensive test of phonological & print processing. Tallahassee, FL: Authors; 2002.
- Lonigan, CJ., Wagner, RK., Torgesen, JK., Rashotte, CA. Test of preschool early literacy (TOPEL). Austin, TX: Pro-Ed; 2007.
- Lonigan, CJ., Wilson, SB. Revised Get Ready to Read! Screening Tool: Psychometrics & normative information. National Center for Learning Disabilities; NY, NY: 2008.
- McMaster KL, Fuchs D, Fuchs LS, Compton DL. Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*. 2005; 71:445–463.
- Muter V, Hulme C, Snowling MJ, Stevenson J. Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology*. 2004; 40:665–681. [PubMed: 15355157]
- Muthén, LK., Muthén, BO. Mplus user's guide. 5. Los Angeles, CA: Muthén & Muthén; 1998–2007.

- O'Connor RE, Bocian K, Beebe-Frankenberger M, Linklater DL. Responsiveness of students with language difficulties to early intervention in reading. *The Journal of Special Education*. 2010; 43:220–235.
- O'Connor RE, Bocian KM, Sanchez V, Beach KD. Access to a responsiveness to intervention model: Does beginning intervention in kindergarten matter? *Journal of Learning Disabilities*. 2012; 47:307–328. [PubMed: 23019070]
- O'Connor RE, Klingner J. Poor responders in RTI. *Theory Into Practice*. 2010; 49:297–304.
- Proctor B, Prevatt F. Agreement among four models used for diagnosing learning disabilities. *Journal of Learning Disabilities*. 2003; 36:459–466. [PubMed: 15497489]
- Reid, DK., Hresko, WP., Hammil, DD. *Test of Early Reading Ability*. 3. Austin, TX: Pro-Ed; 2001. (TERA-3)
- Rodgers B. The identification and prevalence of specific reading retardation. *British Journal of Educational Psychology*. 1983; 53:369–373. [PubMed: 6652037]
- Rutter M, Yule W. The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry*. 1975; 16:181–197. [PubMed: 1158987]
- Schatschneider C, Fletcher JM, Francis DJ, Carlson CD, Foorman BR. Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*. 2004; 96:265–282.
- Schatschneider C, Wagner RK, Crawford EC. The importance of measuring growth in response to intervention models: Testing a core assumption. *Learning and Individual Differences*. 2008; 18:308–315. [PubMed: 22224065]
- Share DL, Silva PA. Language deficits and specific reading retardation: Cause or effect? *International Journal of Language & Communication Disorders*. 1987; 22:219–226.
- Siegel LS. IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities*. 1989; 22:469–478. [PubMed: 2794763]
- Speece DL, Case LP. Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology*. 2001; 93:735–749.
- Speece DL, Case LP, Molloy DE. Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research and Practice*. 2003; 18:147–156.
- Storch SA, Whitehurst GJ. Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental Psychology*. 2002; 38:934–947. [PubMed: 12428705]
- Stuebing KK, Hatcher JM, LeDoux JM, Lyon GR, Shaywitz SE, Shaywitz BA. Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal*. 2002; 39:469–518.
- Tran L, Sanchez T, Arellano B, Swanson HL. A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities*. 2011; 44:283–295. [PubMed: 21521870]
- Torgesen JK. The prevention of reading difficulties. *Journal of School Psychology*. 2002; 40:7–26.
- Torgesen JK. Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice*. 2000; 15:55–64.
- Torgesen JK, Wagner RK, Rashotte CA, Rose E, Lindamood P, Conway T, Garvan C. Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*. 1999; 91:579–593.
- Torgesen JK, Alexander AW, Wagner RK, Rashotte CA, Voeller KK, Conway T. Intensive remedial instruction for children with severe reading disabilities immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*. 2001; 34:33–58. [PubMed: 15497271]
- U.S. Department of Education. *Individuals with disabilities improvement act of 2004*, Pub. L. 108–466. Federal register. 2004; 70:35802–35803.
- U.S. Department of Education. Thirty-Sixth Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2014. 2014. Retrieved from <http://www2.ed.gov/about/reports/annual/osep/2014/parts-b-c/36th-idea-arc.pdf>

- Vaughn S, Fuchs LS. Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice*. 2003; 18:137–146.
- Vaughn S, Linan-Thompson S, Hickman P. Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children*. 2003; 69:391–409.
- Vellutino FR, Scanlon DM, Lyon GR. Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities*. 2000; 33:223–238. [PubMed: 15505962]
- Vellutino F, Scanlon D, Sipay E, Small S, Pratt A, Chen R, Denckla M. Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*. 1996; 88:601–638.
- Vellutino FR, Scanlon DM, Zhang H, Schatschneider C. Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing: An Interdisciplinary Journal*. 2008; 21:437–480.
- Wechsler, D. Wechsler preschool and primary scale of intelligence. 3. San Antonio, TX: Harcourt Assessment; 2002. (WPPSI-III)
- Wiig, EH., Secord, W., Semel, E. Clinical evaluation of language fundamentals–preschool. San Antonio, TX: The Psychological Corporation; 1992.

Highlights

- Agreement among methods of indexing risk for learning disability in preschoolers was low.
- We assessed 1,011 preschoolers in language, phonological awareness, and print knowledge.
- We used chance-corrected affected-status agreement (CCASA) statistic.
- Rates of agreement generally decreased with greater levels of severity.
- These low rates of agreement represent disparity of children identified across methods.

Table 1

Descriptive Statistics for End-of-Year Emergent Literacy Skill and IQ Measures

Measure	Mean	SD	Range Min - Max
Emergent Literacy			
Pre-CTOPPP Definitional Vocabulary ^a	95.93	12.26	55–118
CELF Expressive Language ^b	95.08	15.96	50–146
Pre-CTOPPP Print Knowledge ^c	103.06	13.97	69–128
TERA Alphabet Knowledge ^d	9.59	3.25	3–20
Pre-CTOPPP Elision ^e	10.34	2.86	2–16
Pre-CTOPPP Blending ^f	9.95	2.70	1–14
IQ			
WPPSI Block Design ^g	8.06	3.25	1–18
WPPSI Matrix Reasoning ^h	8.47	3.40	1–18

Note.

^a
n = 1,002.^b
n = 1,006.^c
n = 1,003.^d
n = 1,003.^e
n = 1,001.^f
n = 1,001.^g
n = 1,011.^h
n = 1,011.

Pre-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processing; CELF = Clinical Evaluation of Language Fundamentals, Fourth Edition; TERA = Test of Early Reading Ability; WPPSI = Wechsler Preschool and Primary Scale of Intelligence, Third Edition. Pre-CTOPPP Definitional Vocabulary, CELF Expressive Language, and Pre-CTOPPP Print Knowledge descriptive scores are standard scores and TERA Alphabet Knowledge, Pre-CTOPPP Elision and Blending, and WPPSI Block Design and Matrix Reasoning descriptive scores are scaled score.

Table 2

Percentage of Children Classified as At-Risk for LD using Four Definitions at Three Levels of Severity

	Below 25 th Percentile	Below 10 th Percentile	Below 5 th Percentile
IQ-Achievement Discrepancy			
ULA			
Pre-CTOPPP DV	6.0	4.1	3.5
CELF EL	8.7	6.6	4.4
TERA AL	7.6	6.9	3.3
Pre-CTOPPP PK	3.7	1.9	1.5
Pre-CTOPPP Elision	2.8	2.1	1.7
Pre-CTOPPP Blending	5.8	4.8	4.0
RTI			
LA			
Pre-CTOPPP DV	28.2	11.8	7.6
CELF EL	36.7	17.3	10.5
TERA AL	40.9	30.1	10.5
Pre-CTOPPP PK	22.7	10.3	5.1
Pre-CTOPPP Elision	28.7	18.3	9.8
Pre-CTOPPP Blending	27.9	19.1	12.7
LG			
Pre-CTOPPP DV	26.3	6.6	1.6
CELF EL	23.6	9.7	4.0
TERA AL	27.5	7.3	3.2
Pre-CTOPPP PK	30.2	6.9	1.9
Pre-CTOPPP Elision	25.2	9.2	5.3
Pre-CTOPPP Blending	23.0	11.8	6.7
DD			
Pre-CTOPPP DV	3.7	0.3	0.0
CELF EL	21.4	7.4	2.8
TERA AL	21.9	13.9	1.6
Pre-CTOPPP PK	13.9	3.9	0.8
Pre-CTOPPP Elision	16.9	6.6	3.8
Pre-CTOPPP Blending	17.7	9.7	5.1

Note. *N* = 1,011. ULA = unexpected low achievement; LA = low achievement; LG = low growth; DD = dual discrepancy; Pre-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processing; DV = Definitional Vocabulary; PK = Print Knowledge; CELF = Clinical Evaluation of Language Fundamentals, Fourth Edition; EL = Expressive Language; TERA AL = Test of Early Reading Ability; AL = Alphabet.

Table 3

Affected Status Agreement Rates [and 95% Confidence Intervals] for Classifying Children as At-Risk for LD Using Four Methods at Three Levels of Severity for Single Measures

	Below 25 th Percentile	Below 10 th Percentile	Below 5 th Percentile
ULA vs. LA			
Pre-CTOPPP DV	.17 [.12, .22]	.32 [.24, .41]	.44 [.33, .56]
CELF EL	.18 [.13, .23]	.35 [.27, .42]	.40 [.30, .49]
TERA AL	.12 [.09, .16]	.18 [.13, .23]	.30 [.21, .38]
Pre-CTOPPP PK	.13 [.09, .18]	.17 [.10, .24]	.27 [.14, .39]
Pre-CTOPPP Elision	.07 [.04, .11]	.10 [.05, .14]	.16 [.09, .24]
Pre-CTOPPP Blending	.16 [.12, .21]	.22 [.16, .28]	.29 [.21, .37]
ULA vs. LG			
Pre-CTOPPP DV	-.02 [-.04, .00]	-.01 [-.03, .02]	-.01 [-.01, -.01]
CELF EL	.23 [.17, .28]	.27 [.19, .36]	.15 [.06, .24]
TERA AL	.14 [.09, .18]	.11 [.05, .17]	.07 [.00, .14]
Pre-CTOPPP PK	.04 [.01, .07]	.14 [.06, .23]	.13 [.00, .26]
Pre-CTOPPP Elision	.07 [.04, .11]	.11 [.05, .18]	.17 [.08, .28]
Pre-CTOPPP Blending	.17 [.12, .22]	.23 [.16, .30]	.32 [.22, .42]
ULA vs. DD			
Pre-CTOPPP DV	.09 [.03, .16]	.04 [-.02, .11]	.00 [.00, .00]
CELF EL	.26 [.20, .32]	.35 [.26, .44]	.18 [.08, .29]
TERA AL	.19 [.14, .24]	.14 [.07, .21]	.10 [.01, .20]
Pre-CTOPPP PK	.13 [.07, .19]	.25 [.12, .38]	.21 [.03, .41]
Pre-CTOPPP Elision	.12 [.07, .18]	.16 [.08, .25]	.24 [.11, .37]
Pre-CTOPPP Blending	.23 [.17, .30]	.28 [.20, .37]	.41 [.29, .53]
LA vs. LG			
Pre-CTOPPP DV	-.10 [-.12, -.08]	-.03 [-.05, -.01]	-.01 [-.01, -.01]
CELF EL	.46 [.41, .50]	.33 [.26, .40]	.21 [.14, .29]
TERA AL	.34 [.30, .39]	.13 [.09, .17]	.11 [.05, .17]
Pre-CTOPPP PK	.24 [.19, .29]	.26 [.18, .34]	.12 [.03, .20]
Pre-CTOPPP Elision	.36 [.31, .41]	.27 [.20, .33]	.31 [.22, .40]
Pre-CTOPPP Blending	.46 [.40, .51]	.41 [.35, .48]	.33 [.25, .41]
LA vs. DD			
Pre-CTOPPP DV	.10 [.06, .14]	.02 [.00, .05]	.00 [.00, .00]
CELF EL	.51 [.46, .56]	.39 [.32, .46]	.25 [.16, .33]
TERA AL	.44 [.40, .49]	.13 [.08, .17]	.14 [.07, .21]
Pre-CTOPPP PK	.57 [.51, .63]	.36 [.27, .45]	.15 [.05, .25]
Pre-CTOPPP Elision	.53 [.48, .59]	.33 [.26, .39]	.37 [.27, .45]
Pre-CTOPPP Blending	.59 [.53, .64]	.47 [.40, .54]	.38 [.30, .47]
LG vs. DD			
Pre-CTOPPP DV	.11 [.07, .15]	.04 [.00, .09]	.00 [.00, .00]
CELF EL	.89 [.86, .93]	.75 [.67, .84]	.69 [.55, .84]

	Below 25th Percentile	Below 10th Percentile	Below 5th Percentile
TERA AL	.76 [.72, .81]	.79 [.69, .88]	.49 [.32, .67]
Pre-CTOPPP PK	.39 [.34, .45]	.56 [.44, .67]	.45 [.23, .66]
Pre-CTOPPP Elision	.63 [.57, .69]	.70 [.61, .80]	.71 [.59, .83]
Pre-CTOPPP Blending	.74 [.69, .80]	.81 [.74, .88]	.75 [.65, .85]

Note. $N = 1,011$. ULA = unexpected low achievement; LA = low achievement; LG = low growth; DD = dual discrepancy; Pre-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processing; DV = Definitional Vocabulary; PK = Print Knowledge; CELF = Clinical Evaluation of Language Fundamentals, Fourth Edition; EL = Expressive Language; TERA = Test of Early Reading Ability; AL = Alphabet; PCTOPP = Preschool Comprehensive Test of Phonological Processing. Affected-status agreement values were corrected for chance. Bolded values are those that were not significantly better than chance.

Table 4

Percentage of Children Classified as At-Risk for LD Based on Pervasive Deficits using Four Methods at Three Levels of Severity

	Below 25 th Percentile	Below 10 th Percentile	Below 5 th Percentile
Pervasive ULA			
Language ^a	3.1	2.3	1.6
Print Knowledge ^b	2.9	1.3	0.8
Phonological Awareness ^c	0.9	0.6	0.4
Pervasive LA			
Language ^a	20.6	8.7	4.5
Print Knowledge ^b	21.3	9.7	3.2
Phonological Awareness ^c	13.8	7.3	2.7
Pervasive LG			
Language ^d	2.3	0.2	0.0
Print Knowledge ^d	11.8	1.2	0.1
Phonological Awareness ^a	7.7	1.7	0.8
Pervasive DD			
Language ^d	0.7	0.1	0.0
Print Knowledge ^d	8.0	0.9	0.1
Phonological Awareness ^a	4.3	1.3	0.6

Note.

^a *n* = 998.

^b *n* = 999.

^c *n* = 1,001.

^d *n* = 996.

ULA = unexpected low achievement; LA = low achievement; LG = low growth; DD = dual discrepancy.

Table 5

Affected Status Agreement Rates [and 95% Confidence Intervals] for Classifying Children as At-Risk for LD Using Four Definitions at Three Levels of Severity Based on Pervasive Deficits

	Below 25 th Percentile	Below 10 th Percentile	Below 5 th Percentile
ULA vs. LA			
Language ^a	.13 [.10, .20]	.25 [.17, .36]	.35 [.22, .50]
Print Knowledge ^b	.11 [.09, .18]	.12 [.07, .20]	.24 [.10, .40]
Phonological Awareness ^c	.06 [.02, .11]	.08 [.02, .14]	.14 [.01, .28]
ULA vs. LG			
Language ^d	.01 [–.02, .06]	.00 [.00, .00]	.00 [.00, .00]
Print Knowledge ^d	.07 [.05, .15]	.13 [–.01, .28]	.12 [–.10, .35]
Phonological Awareness ^a	.07 [.01, .13]	.15 [–.01, .31]	.33 [.02, .64]
ULA vs. DD			
Language ^d	.02 [–.02, .08]	.00 [.00, .00]	.00 [.00, .00]
Print Knowledge ^d	.12 [.06, .20]	.16 [–.01, .32]	.12 [–.10, .35]
Phonological Awareness ^a	.12 [.03, .23]	.18 [.00, .38]	.43 [.06, .80]
LA vs. LG			
Language ^d	.01 [–.01, .03]	.00 [–.02, .02]	.00 [.00, .00]
Print Knowledge ^d	.26 [.20, .32]	.08 [.03, .15]	.03 [–.03, .09]
Phonological Awareness ^a	.21 [.14, .27]	.15 [.07, .24]	.20 [.05, .35]
LA vs. DD			
Language ^d	.03 [.00, .05]	.01 [–.01, .03]	.00 [.00, .00]
Print Knowledge ^d	.33 [.27, .40]	.08 [.03, .14]	.03 [–.03, .09]
Phonological Awareness ^a	.29 [.21, .36]	.17 [.08, .26]	.22 [.06, .38]
DD vs. LG			
Language ^d	.30 [.12, .49]	.50 [–.19, 1.19]	.00 [.00, .00]
Print Knowledge ^d	.66 [.59, .76]	.75 [.50, 1.00]	1.00 [1.00, 1.00]
Phonological Awareness ^a	.54 [.45, .67]	.76 [.56, .97]	.75 [.45, 1.05]

Note.

^a
n = 998.

^b
n = 999.

^c
n = 1,001.

^d
n = 996.

ULA = unexpected low achievement; LA = low achievement; LG = low growth; DD = dual discrepancy; Affected-status agreement values were corrected for chance. Bolded values are those that were not significantly better than chance.